FAST BAYESIAN NMF ALGORITHMS ENFORCING HARMONICITY AND TEMPORAL CONTINUITY IN POLYPHONIC MUSIC TRANSCRIPTION

Nancy Bertin, Roland Badeau[*]

CNRS LTCI, TELECOM ParisTech (ENST)
75634 Paris Cedex 13, FRANCE
{nancy.bertin,roland.badeau}@telecom-paristech.fr

Emmanuel Vincent

METISS Project, IRISA-INRIA
35 042 Rennes Cedex, FRANCE
emmanuel.vincent@irisa.fr

ABSTRACT

This article presents theoretical and experimental results about constrained non-negative matrix factorization (NMF) in a Bayesian framework, enforcing both spectral harmonicity and temporal continuity. We exhibit fast multiplicative update rules to perform the decomposition, which are then applied to perform polyphonic piano music transcription. This approach is shown to outperform other standard NMF-based transcription systems, providing a meaningful mid-level representation of the data.

Index Terms— Non-negative matrix factorization (NMF), music transcription, audio source separation, Bayesian regression.

1. INTRODUCTION

Non-negative matrix factorization (NMF) is a powerful, unsupervised decomposition technique allowing the representation of two-dimensional non-negative data as a linear combination of meaningful elements in a basis. NMF has been widely and successfully used to process audio signals, including various tasks such as monaural sound source separation [1] and music transcription [2]. In the latter case, a time-frequency representation of the signal is factored as the product between a basis (or dictionary) of pseudo-spectra and a matrix (decomposition) of time-varying gains. When obtained from harmonic instruments sounds, the basis is shown to partially retain harmonic components, with a pitched structure, that can be interpreted as musical notes, while the decomposition gives information about the onset and offset times of the associated notes. This interpretability is often observed in practice, which is certainly one of the reasons for NMF's popularity; but it is not always as satisfying as expected. In this paper we focus on a Bayesian approach to NMF that allows to enforce harmonicity of the dictionary components (a desired property for music transcription) and temporal smoothness of the decomposition. This approach was first investigated in [3], where we introduced EM-based algorithms to perform the decomposition. In order to reduce the computational burden, we herein propose an alternative approach inspired by multiplicative heuristics [4]. The paper is organized as follows. Section 2 addresses the baseline NMF model, the constraint of harmonicity, and Bayesian approaches to NMF. Our statistical model, and fast multiplicative update rules enforcing harmonicity and smoothness are presented in section 3. Section 4 is devoted to experimental results in the context of music transcription. Conclusion and perspectives are drawn in section 5.

[*]The research leading to this paper was supported by the French GIP ANR under contract ANR-06-JCJC-0027-01, DESAM, and by the Quaero Programme, funded by OSEO, French State agency for innovation.

2. CONSTRAINED NON-NEGATIVE MATRIX FACTORIZATION

Throughout the paper, matrices are denoted by straight bold letters, for instance $\mathbf{W} = (w_{fk})$, $\mathbf{H} = (h_{kn})$ and $\mathbf{V} = (v_{fn})$. Lowercase bold letters denote column vectors, such that $\mathbf{w}_k = (w_{1k} \dots w_{Fk})^T$, while lowercase plain letters with a single index denote rows, such that $\mathbf{H} = (h_1^T \dots h_K^T)^T$. The binary operators $\triangleq$ and $\overset{c}{=}$ denote definitions and equality up to an additive constant.

2.1. Baseline model and algorithms

Out of any applicative context, the NMF problem is expressed as follows: given a matrix $\mathbf{V}$ of dimensions $F \times N$ with non-negative entries, NMF is the problem of finding a factorization $\hat{\mathbf{V}} \triangleq \mathbf{WH} \approx \mathbf{V}$, where $\mathbf{W}$ and $\mathbf{H}$ are non-negative matrices of dimensions $F \times K$ and $K \times N$, respectively. K is usually chosen such that $FK + KN \ll FN$, hence reducing the data dimension. In typical audio applications, the matrix $\mathbf{V}$ is often the magnitude or power spectrogram, f denoting the frequency bin and n the time frame. This factorization is obtained by minimizing a cost function defined by

$$D(\mathbf{V}|\hat{\mathbf{V}}) = \sum_{f=1}^{F} \sum_{n=1}^{N} d(v_{fn}|\hat{v}_{fn}) \tag{1}$$

where $d(a|b)$ is a function of two scalar variables. d is typically non-negative and takes value zero if and only if (iff) $a = b$. The most popular cost functions for NMF are the Euclidean (EUC) distance and the generalized Kullback-Leibler (KL) divergence, which were particularly popularized (as NMF itself) by Lee and Seung, see, *e.g.*, [4]. They described multiplicative update rules under which $D(\mathbf{V}|\mathbf{WH})$ is shown to be non-increasing, while ensuring non-negativity of $\mathbf{W}$ and $\mathbf{H}$. These rules follow a simple heuristics, which can be seen as a gradient descent algorithm with an appropriate choice of the descent step[1]. They are obtained by expressing the partial derivatives of the cost function ∇D as the difference of two positive terms $\nabla^+ D$ and $\nabla^- D$:

$$\begin{cases} w_{fk} \leftarrow w_{fk} \times \dfrac{\nabla^-_{w_{fk}} D(\mathbf{V}|\mathbf{WH})}{\nabla^+_{w_{fk}} D(\mathbf{V}|\mathbf{WH})} \\ h_{kn} \leftarrow h_{kn} \times \dfrac{\nabla^-_{h_{kn}} D(\mathbf{V}|\mathbf{WH})}{\nabla^+_{h_{kn}} D(\mathbf{V}|\mathbf{WH})} \end{cases} \tag{2}$$

[1]For some choices of d, like EUC/KL, monotonicity of the criterion under these rules is proved [4], but convergence is not guaranteed in general.

2.2. Constrained approaches

In standard NMF, the only constraint is the element-wise non-negativity of all matrices. All other properties of the decomposition come as uncontrolled side-effects. It sounds thus natural to improve this potential by adding explicit constraints to the factorization problem, in order to enhance and control desired properties.

2.2.1. Deterministic constraints

Musical notes, excluding transients, are pseudo-periodic. Their spectra consist in regularly spaced frequency peaks. As we wish to use NMF to identify musical notes in a polyphonic recording, we expect that elements in the basis $\mathbf{W}$ follow this harmonic shape. In [5], we proposed an alternative model enforcing harmonicity. We impose the basis components to be expressed as the linear combination of fixed narrow-band harmonic spectra (patterns):

$$w_{fk} = \sum_{m=1}^{M} e_{mk} P_{km}(f). \tag{3}$$

For a given component index k, all the patterns $P_{km}(f)$ share the same pitch (fundamental frequency f_0); they are defined by summation of the spectra of a few adjacent individual partials at harmonic frequencies of f_0, scaled by the spectral shape of sub-band k. This spectral envelope is chosen according to perceptual modeling[2]. $\mathbf{E}$ can be interpreted as global frequency envelope coefficients for one component $\mathbf{w}_k$. The coefficients e_{mk} are learned by NMF as well as the decomposition coefficients $\mathbf{H}$. Update rules are obtained by minimizing the same cost function as in baseline NMF, except that it is minimized with respect to (wrt) $\mathbf{E}$ and $\mathbf{H}$ rather than $\mathbf{W}$ and $\mathbf{H}$.

2.2.2. Statistical constraints

Another way to induce properties in NMF is to switch to a statistical framework and introduce adequate prior distributions. Let us consider a complex-valued time-frequency representation $\mathbf{X}$ of the signal, and the following model: $\forall n = 1, ..., N$,

$$\mathbf{x}_n = \sum_{k=1}^{K} \mathbf{c}_{kn} \in \mathbb{C}^F \tag{4}$$

where latent variables $\mathbf{c}_{kn}$ are independent and follow a multivariate complex Gaussian distribution[3]: $\mathbf{c}_{kn} \sim \mathcal{N}(0, h_{kn}\mathrm{diag}(\mathbf{w}_k))$. The estimation of the parameters $\boldsymbol{\theta} = \{\mathbf{W}, \mathbf{H}\}$ in a maximum likelihood (ML) sense is performed by maximizing the criterion

$$C_{ML}(\boldsymbol{\theta}) \triangleq \log p(\mathbf{X}|\boldsymbol{\theta}). \tag{5}$$

Then it is easily proved that $C_{ML}(\boldsymbol{\theta}) \stackrel{c}{=} -D(\mathbf{V}|\hat{\mathbf{V}})$, where $\mathbf{V} = |\mathbf{X}|^2$, and the cost function d is the Itakura-Saito (IS) divergence:

$$d_{IS}(a|b) = \frac{a}{b} - \log\frac{a}{b} - 1. \tag{6}$$

Thus ML estimation is equivalent to solving the NMF problem $\mathbf{V} \approx \mathbf{WH}$ (see [6] for a full study and justification of this model).

[2]Figure 1 in [5] illustrates the narrow-band spectra and spectral envelope coefficients for one note, and the corresponding harmonic spectrum.

[3]Gaussian distribution: $\mathcal{N}(\mathbf{u}|\boldsymbol{\mu}, \boldsymbol{\Sigma})=\frac{1}{\det(\pi\boldsymbol{\Sigma})} e^{-(\mathbf{u}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{u}-\boldsymbol{\mu})}$, where the symbol ^{H} denotes the conjugate transpose.

This approach offers the possibility to switch to maximum a posteriori (MAP) estimation, thanks to Bayes rule:

$$p(\mathbf{W}, \mathbf{H}|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{W}, \mathbf{H})p(\mathbf{W})p(\mathbf{H})}{p(\mathbf{X})} \tag{7}$$

Thus, choosing adequate prior distributions $p(\mathbf{W})$ and $p(\mathbf{H})$ is a way to induce desired properties in the decomposition. Below, we propose to combine this framework and the previous model (3) to enforce both harmonicity in $\mathbf{W}$ and smoothness in $\mathbf{H}$, which are desired properties of the NMF of musical signals.

3. PROPOSED ALGORITHM

In [3], the estimation is performed by means of an EM-like algorithm, whose local convergence is guaranteed, but which remains slow compared to multiplicative gradient descent approaches. In this section, we thus propose the direct optimization of the criterion by an adaptation of the multiplicative heuristics (2).

3.1. Probabilistic harmonic model

We now describe multiplicative update rules for ML estimation of the parameters $\boldsymbol{\theta} = \{\mathbf{E}, \mathbf{H}\}$. By substituting constraint (3) into model (4), the cost function to be minimized becomes

$$D_{IS}(\mathbf{V}|\mathbf{WH}) = \sum_{f=1}^{F}\sum_{n=1}^{N} d_{IS}\left(v_{fn} \,|\, \sum_{k=1}^{K}\sum_{m=1}^{M} h_{kn} e_{mk} P_{km}(f)\right) \tag{8}$$

We compute its derivative wrt h_{kn}, which is expressed as the difference of two positive terms:

$$\nabla_{h_{kn}} D_{IS}(\mathbf{V}|\mathbf{WH}) = \sum_{f=1}^{F} \frac{w_{fk}}{\hat{v}_{fn}} - \sum_{f=1}^{F} \frac{v_{fn} w_{fk}}{\hat{v}_{fn}^2} \tag{9}$$

where $\hat{v}_{fn} = \sum_{k'=1}^{K} w_{fk'} h_{k'n} = \sum_{k'=1}^{K}\sum_{m'=1}^{M} e_{m'k'} P_{k'm'}(f) h_{k'n}$. The derivative wrt e_{mk} fits in the same scheme:

$$\nabla_{e_{mk}} D_{IS}(\mathbf{V}|\mathbf{WH}) = \sum_{f=1}^{F}\sum_{n=1}^{N} \frac{h_{kn} P_{km}(f)}{\hat{v}_{fn}} - \sum_{f=1}^{F}\sum_{n=1}^{N} \frac{v_{fn} h_{kn} P_{km}(f)}{\hat{v}_{fn}^2} \tag{10}$$

The update rules are derived from the heuristics (2) and can be written:

$$h_{kn} \leftarrow h_{kn} \times \frac{\sum_{f=1}^{F} v_{fn} w_{fk}/\hat{v}_{fn}^2}{\sum_{f=1}^{F} w_{fk}/\hat{v}_{fn}} \tag{11}$$

$$e_{mk} \leftarrow e_{mk} \times \frac{\sum_{f=1}^{F}\sum_{n=1}^{N} v_{fn} h_{kn} P_{km}(f)/\hat{v}_{fn}^2}{\sum_{f=1}^{F}\sum_{n=1}^{N} h_{kn} P_{km}(f)/\hat{v}_{fn}} \tag{12}$$

This algorithm will be referred to as "H-NMF/MU".

3.2. Enforcing temporal smoothness

The maximum likelihood estimation of $\mathbf{E}$ and $\mathbf{H}$ opens the possibility of constraining NMF solutions by including priors on the parameters. As in [6], this framework is exploited to enforce temporal smoothness over the rows of $\mathbf{H}$. We provide a priori information on $\boldsymbol{\theta}$, expressed as a prior distribution $p(\boldsymbol{\theta})$. Thanks to Bayes rule (7), we get a maximum a posteriori (MAP) estimator by maximizing the following criterion:

$$\begin{aligned} C_{MAP}(\boldsymbol{\theta}) &\triangleq \log p(\boldsymbol{\theta}|\mathbf{X}) \\ &\stackrel{c}{=} C_{ML}(\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) \end{aligned}$$

We choose the Markov chain prior structure proposed in [6]:

$$p(h_k) = p(h_{k1}) \prod_{n=2}^{N} p(h_{kn}|h_{k(n-1)}) \tag{13}$$

where $p(h_{kn}|h_{k(n-1)})$ reaches its maximum at $h_{k(n-1)}$, thus favoring a slow variation of h_k in time. We propose to choose

$$p(h_{kn}|h_{k(n-1)}) = \mathcal{IG}(h_{kn}|\alpha_k, (\alpha_k+1)h_{k(n-1)}) \tag{14}$$

where $\mathcal{IG}(u|\alpha,\beta)$ is the inverse-Gamma distribution[4] with mode $\beta/(\alpha+1)$ and the initial distribution $p(h_{k1})$ is Jeffrey's non-informative prior: : $p(h_{k1}) \propto 1/h_{k1}$. Parameters α_k are here arbitrarily fixed, depending on the desired degree of smoothness (the higher α_k, the smoother h_k), but we could consider in future work the possibility to learn them as well. We do not put here any prior on $\mathbf{E}$. In order to compute NMF with both harmonicity and smoothness constraints, we directly use the update rules (2) with

$$-C^{MAP}(\boldsymbol{\theta}) = D_{IS}(\mathbf{V}|\mathbf{WH}) - \sum_{k=1}^{K} \log(p(h_k)),$$

where the contribution of the prior can be seen as a penalty term. Updates for $\mathbf{E}$ are unchanged and we obtain new update rules for $\mathbf{H}$. However, first simulation experiments showed that under this update scheme, the criterion was not always monotonically decreasing. Then, we propose to raise the ratio in (2) to a certain power $\eta \in]0,1[$, whose role is similar to the step size in usual gradient descents. We then obtain the following update rules: for $n = 2 \ldots N-1$,

$$h_{kn} \leftarrow h_{kn} \times \left(\frac{\sum_{f=1}^{F} \frac{v_{fn} w_{fk}}{\hat{v}_{fn}^2} + \frac{(\alpha_k+1)h_{k,n-1}}{h_{kn}^2}}{\sum_{f=1}^{F} \frac{w_{fk}}{\hat{v}_{fn}} + \frac{1}{h_{kn}} + \frac{\alpha_k+1}{h_{k,n+1}}} \right)^{\eta} \tag{15}$$

Similar updates are determined for the boundaries of the Markov chain ($n = 1$ and $n = N$). This algorithm will be referred to as "HS-NMF/MU".

4. APPLICATION TO MUSIC TRANSCRIPTION

Music transcription consists in converting a raw music signal into a symbolic representation of the music within: for instance a score, or a MIDI file. Here, we focus on information strictly related to musical notes, *i.e.* pitch, onset and offset time. Various methods have been proposed to address the transcription issue, including neural network modeling [7], or parametric signal modeling and HMM tracking [8]. We propose here to assess the efficiency of Bayesian harmonic and smooth NMF for this task.

4.1. Experimental setup

4.1.1. Database

To evaluate and quantify transcription performance, we need a set of polyphonic music pieces with accurate MIDI references. Here, we use a subset of Valentin Emiya's *MAPS* (MIDI-Aligned Piano Sounds) database, which include, among others, pieces from the piano repertoire either recorded on an upright DisKlavier or produced by high quality software synthesis, with associated MIDI groundtruth. From this very complete database, we selected two subsets to evaluate our algorithms: a synthetic subset, produced by Native Instruments' Akoustik Piano, and a real audio subset, recorded at Tlcom ParisTech on a Yamaha Mark III (upright DisKlavier). Each subset is composed of 30 pieces of 30 seconds each (original pieces from *MAPS* were truncated).

4.1.2. Structure of NMF-based transcription

All NMF-based transcription systems used here follow the same workflow:

1. Computation of a time-frequency representation, $\mathbf{V}$;
2. Factorization $\mathbf{V} \approx \mathbf{WH}$;
3. Attribution of a MIDI pitch to each basis spectrum $\mathbf{w}_k$;
4. Onset/offset detection applied to each time envelope h_k.

In [5], it is observed that using a nonlinear frequency scale results in a representation of smaller size, with better temporal resolution in the higher frequency range, than the usual Short-Time Fourier Transform (STFT), while preserving the subsequent transcription performance. We then pass the signal through a filterbank of 257 sinusoidally modulated Hanning windows with frequencies linearly spaced between 5 Hz and 10.8 kHz on the Equivalent Rectangular Bandwidth (ERB) scale. We finally split each sub-band into disjoint 23 ms time frames and compute the power within each frame. Pitch estimation of basis spectra is superfluous in harmonically constrained NMF, since each basis component can be labeled from the beginning with the pitch of the patterns $P_{km}(f)$ used to initialize it. For NMF without harmonicity constraint, pitch identification is performed on each column of $\mathbf{W}$ by the harmonic comb-based technique used in [5]. Note onsets and offsets are determined by a simple threshold-based detection, followed by a minimum-duration pruning, see [5]. The detection threshold is denoted by A_{dB} and expressed in dB wrt the maximum of $\mathbf{H}$.

4.1.3. Evaluation

Transcription performance is quantitatively evaluated according to usual information retrieval scores. **Precision rate** ($\mathcal{P}$) is the proportion of correct notes among all transcribed notes. **Recall rate** ($\mathcal{R}$) is the proportion of notes from the MIDI reference which are correctly transcribed. **F-measure** ($\mathcal{F}$) aggregates the two former criteria in one unique score and is defined as $\mathcal{F} = 2\mathcal{PR}/(\mathcal{P}+\mathcal{R})$. A transcribed note is considered as correct if its pitch is identical to the ground truth, and its onset time is within 50ms of the ground

[4] Inverse-Gamma: $\mathcal{IG}(u|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} u^{-(\alpha+1)} \exp(-\frac{\beta}{u})$, $u \geq 0$.

truth, according to community standards. The original algorithms previously proposed are compared to several state-of-the-art algorithms listed in Table 1.

Abbr.	Description	Ref.
NMF/MU	Baseline NMF minimizing IS divergence Multiplicative update rules	[6]
S-NMF	SAGE algorithm for NMF With smoothness constraint on $\mathbf{H}$	[6]
HS-NMF	SAGE algorithm for NMF With both harmonicity and smoothness	[3]
Virtanen'07	Multiplicative NMF With temporal continuity constraint	[1]
Marolt'04	Neural network based transcription	[7]
Emiya'08	Joint multipitch estimation and HMM tracking	[8]

Table 1: Reference algorithms.

Virtanen'07, Emiya'08 and NMF/MU are run from their author's implementation, which they nicely shared, and Marolt'04 is run from the SONIC software, distributed by its author. The other algorithms were implemented by the authors. The order K is set to 88 (the number of keys on a piano) for all NMF-based algorithms. HS-NMF and S-NMF are initialized with 10 iterations of the corresponding multiplicative algorithm (H-NMF/MU and NMF/MU respectively). Note detection thresholds A_{dB} are manually tuned (and reported in Tables 2 and 3) by maximizing the average F-measure on each dataset. The minimum duration for a transcribed note is fixed to 50ms. The step size η is set to 0.5.

4.2. Results

Algorithm	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	A_{dB}
NMF/MU	63.4	56.1	54.9	-62
H-NMF/MU	58.7	59.1	52.4	-33
S-NMF	62.4	43.3	49.5	-51
Virtanen'07	55.9	56.4	53.6	-22
HS-NMF	65.8	64.5	60.7	-38
HS-NMF/MU	**78.5**	**62.6**	**67.0**	-42
Marolt'04	83.5	70.1	75.8	-
Emiya'08	77.3	61.6	67.7	-

Table 2: Transcription scores on synthetic data.

Algorithm	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	A_{dB}
NMF/MU	43.3	43.4	40.8	-60
H-NMF/MU	43.0	42.7	41.3	-30
S-NMF	46.2	32.0	36.6	-49
Virtanen'07	34.2	34.8	33.6	-21
HS-NMF	46.6	45.3	45.0	-32
HS-NMF/MU	**50.6**	**42.7**	**45.0**	-35
Marolt'04	63.7	53.6	58.0	-
Emiya'08	66.1	45.5	52.9	-

Table 3: Transcription scores on real audio data.

Tables 2 and 3 report the transcription performance of tested algorithms on the synthetic and recorded datasets respectively. HS-NMF/MU outperforms other NMF-based algorithms in both cases and is competitive with Emiya'08, but remains less performant than SONIC software. The smoothness constraint used alone seems detrimental to transcription performance, may it be implemented by a multiplicative algorithm (Virtanen'07) or by a Bayesian algorithm (S-NMF), but improves the performance of harmonically constrained NMF (H-NMF vs. HS-NMF). Considering that Emiya'04 and Marolt'04 are supervised systems, specifically tuned for piano music, these results show that our unsupervised, generic algorithm is competitive with the state-of-the-art. The computational cost of one iteration of HS-NMF/MU is comparable to HS-NMF, but the multiplicative algorithm converges 10 to 20 times faster (in number of iterations and runtime) than its EM counterpart, despite the slowing down induced by the step size η and the speed up brought by multiplicative initialization of HS-NMF, thus validating our approach in terms of computational cost.

5. CONCLUSION AND PERSPECTIVES

We proposed an original model for including harmonicity and temporal smoothness constraints in non-negative matrix factorization of time-frequency representations, in a unified framework. The multiplicative update rules we propose are derived from a Bayesian framework like our previous EM-like algorithms [3], but they converge faster. They outperform other benchmarked NMF approaches in a task of polyphonic music transcription, evaluated on a realistic music database. Possible improvements include a refinement of the temporal prior, which suits for modeling the sustain and decay parts of the note, but disfavor attacks and silences.

6. REFERENCES

[1] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, mar 2007.

[2] P. Smaragdis and J. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03)*, New Paltz, New York, USA, October 19-22 2003, pp. 177–180.

[3] N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription." *IEEE Trans. on Audio, Speech and Language Processing*, 2008, accepted with minor modifications.

[4] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[5] E. Vincent, N. Bertin, and R. Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP'08)*, Las Vegas, Nevada, USA, March 30 - April 4, 2008, pp. 109–112.

[6] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[7] M. Marolt, "A connectionist approach to automatic transcription of polyphonic piano music," *IEEE Trans. on Multimedia*, vol. 6, no. 3, pp. 439–449, June 2004.

[8] V. Emiya, R. Badeau, and B. David, "Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches," in *Proc. Eur. Conf. Sig. Proces. (EUSIPCO)*, Lausanne, Suisse, August 25-29 2008.